

Tartu Ülikool  
Loodus- ja tehnoloogiateaduskond  
Keemia Instituut

Hardi Haljas

**Inimese papilloomiviiruse HPV6 E1 helikaasi ATPaasse aktiivsuse inhibeerimise  
modelleerimine QSAR meetodiga**

Bakalaureusetöö

Juhendajad: Kaido Tamm, *Ph. D*  
Lauri Sikk, *Ph. D*

Tartu 2013

## Sisukord

Sisukord.....	2
Kasutatud lühendid.....	3
1. Sissejuhatus.....	4
2. Kirjanduse ülevaade.....	5
2.1 Papilloomiviiruste klassifikatsioon.....	5
2.2 Genoomne ülesehitus ja kodeeritavad valgud .....	5
2.3 Papilloomiviiruse elutsükkel .....	7
2.4 Papilloomiviiruse DNA replikatsioon .....	8
2.5 Papilloomiviiruste vastane ravi .....	9
2.6 HPV valgud ravimite sihtmärkidena .....	10
3. Arvutuslik osa .....	12
3.1 Kvantitatiivse struktuur-aktiivsus sõltuvuse (QSAR) meetod.....	12
3.2 Molekulaardeskriptorid .....	12
4. Andmekomplekt ja metoodika.....	14
4.1 Andmekomplekt .....	14
4.2 Metoodika.....	14
5. Tulemused ja analüüs.....	16
5.1 Mudeli valideerimine.....	19
5.2 Mudeli kasutusala .....	20
6. Kokkuvõte.....	21
7. Summary .....	22
8. Kasutatud kirjandus .....	23
Lisad .....	25

## Kasutatud lühendid

AM1 – *Austin Model 1*

ATP – adenosinotriphosfaat (*adenosine triphosphate*)

<sup>2</sup>BIC – sidemeinformatsioonisisalduse indeks (*bonding information content (order 2)*)

BMLR – parim multilineaarne regressioon (*best multilinear regression*)

CPSA – laetud osapindala (*Charged Partial Surface Area*)

DBD – DNA sidumisdomään (*DNA binding domain*)

E1BS – Valgu E1 sidumiskoht (*E1 binding site*)

E2BS – Valgu E2 sidumiskoht (*E2 binding site*)

FPSA – positiivselt laetud osapindala (*fractional positive surface area*)

HDSA – vesiniksideme doonori osapindala (*hydrogen donor surface area*)

HOMO – kõrgeim okupeeritud molekulaarorbitaal (*highest occupied molecular orbital*)

HPV – inimese papilloomiviirus (*human papilloma virus*)

IC<sub>50</sub> – kontsentratsioon, mille juures saavutatakse pool maksimaalsest inhibeerimisest (*half maximal inhibitory concentration*)

LCR – pikk kontrollpiirkond (*long control region*)

LUMO – madalaim mitteokupeeritud molekulaarorbitaal (*lowest unoccupied molecular orbital*)

MOPAC – *Molecular Orbital PAC*kage

ori – replikatsiooni alguspunkt (*origin of replication*)

QSAR – kvantitatiivne struktuur-aktiivsus sõltuvus (*quantitative structure-activity relationship*)

TAD – transaktivatsiooni domään (*transactivation domain*)

URR – ülesvoolu paiknev regulatsioonipiirkond (*upstream regulatory region*)

## 1. Sissejuhatus

Papilloomiviirused võivad põhjustada erinevaid haiguseid alates tavalistest soolatüügastest kuni eluohtlike pahaloomuliste kasvajateni. Üks tõsisemaid inimese papilloomiviiruse (HPV) poolt põhjustatavaid haiguseid on emakakaelavähk. Mõeldes peamiselt emakaelavähiennetusele, on välja töötatud seda põhjustavate nn kõrge riskiga HPV tüüpide (nt HPV 16 ja 18) vastased vaktsiinid. Euroopa Majanduspiirkonnas on väljastatud müügiloa kahevalentsele vaktsiinile (HPV tüüpide 16 ja 18 nakkuse ennetamiseks) ja neljavalentsetele vaktsiinidele (HPV tüübid 6, 11, 16, 18) [1, 2]. Madala riskiga HPV tüüpidest on levinuimad 6 ja 11. Nende põhjustatud on enamus anogenitaalpiirkonna tüügaste juhtudest. Vaktsiinid omavad vaid ennetavat mõju teatud HPV tüüpidega nakatumise eest ja ei ravi juba olemasolevat haigust. Hetkel puudub efektiivne viirusevastane ravim HPV põhjustatud haiguste raviks, olemasolevad ravimeetodid on ebatõhusad või neil esineb muid puuduseid, nt kõrvaltoimed või haiguse taastekkimine) [2].

Papilloomiviiruste intensiivse uurimise tõttu on tema paljunemise ja elutsükli kohta olemas üsna detailsed teadmised. See omakorda on andnud võimaluse otsida spetsiifiliselt HPV vastu suunatud viirusevastast ravimit. Üks potentsiaalne sihtmärk papilloomiviiruse paljunemisprotsessi pärssimisel on papilloomiviiruse ainsa ensümaatilist aktiivsust omava ning DNA replikatsioonist osa võtva valgu inhibeerimine madala molekulmassiga ühenditega.

Käesoleva töö teoreetilises osas antakse kirjanduse alusel põgus ülevaade papilloomiviirustest, keskendudes peamiselt nende elutsükli, DNA replikatsioonile ja replikatsiooniprotsessis osalevatele valkudele. Praktilises osas arendati kvantitatiivne struktuur-aktiivsus sõltuvus (QSAR) mudel HPV tüübi 6 E1 valgu ATPaasse aktiivsuse inhibeerimise kirjeldamiseks. Mudeli loomisel kasutatud eksperimentaalsed andmed pärinevad Faucheri ja tema kaastöötajate poolt 2004. aastal avaldatud artiklist [3].

## **2. Kirjanduse ülevaade**

### **2.1 Papilloomiviiruste klassifikatsioon**

Papilloomiviirused on DNA-viiruste sugukond, mis nakatavad inimeste ja loomade epiteelrakke. Viirused on peremeesorganismide ja epiteelrakkude suhtes väga spetsiifilised. Kehtiva taksonoomia järgi hõlmab papilloomiviiruste sugukond *Papillomaviridae* 30 perekonda, mis jagunevad 69 liigiks, mis omakorda jagunevad tüüpideks [4]. Praeguseks ajaks on teada üle 240 papilloomiviiruse tüübi, nendest 155 on HPV tüübid [5]. Papilloomiviiruste tüüpe eristatakse genoomi kõige rohkem konserveerunud lugemisraami L1 nukleotiidide järjestuse võrdlemisel. Papilloomiviiruste genotüüpe peetakse eraldiseisvateks tüüpideks, kui L1 lugemisraami nukleotiididest kattub vähem kui 90% [6].

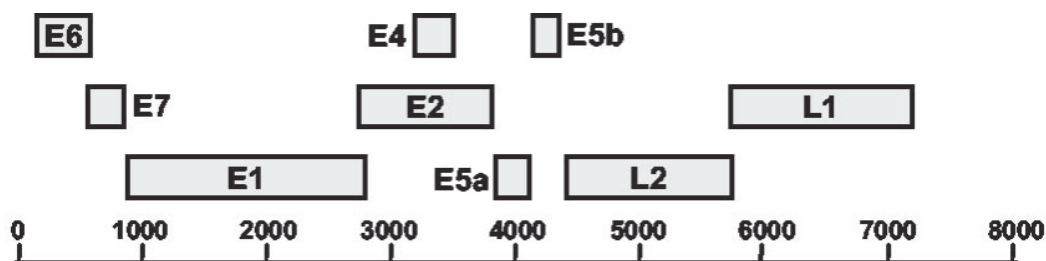
HPV-d kuuluvad eespool mainitud 30-st perekonnast 5 hulka – alfa-, beeta-, gamma-, müü- ja nüüpapilloomiviirused. Gamma-, müü- ja nüüpapilloomiviiruste perekondi seostatakse peamiselt healoomuliste nahapapilloomidega, beetapapilloomiviiruste perekonda pahaloormuliste nahapapilloomidega ja alfa-papilloomiviiruste perekonda kuuluvaid alatüüpe seostatakse nii hea- kui pahaloormuliste limaskestast ja nahapapilloomidega [6].

HPV-d võivad põhjustada naha ja limaskestast epiteelkoe hea- ja pahaloormulisi haiguseid. Ohtlikkuse ehk pahaloormulise kasvaja tekitamise võime alusel jaotatakse papilloomiviirused kõrge ja madala onkogeense riskiga tüüpideks. Tuntuimad kõrge riskiga tüübid on näiteks HPV16, HPV18 ja HPV31, madala riskiga tüüpidest on kõige sagedamini esinevad HPV6 ja HPV11. HPV6 ja HPV11 seostatakse tavaliselt healoomuliste kasvajatega. Nende poolt on põhjustatud näiteks enamus anogenitaalpiirkonna tüügaste juhtudest, samuti põhjustavad nad hingamisteede healoomulist haigust larüngeaalne papillomatoos. Harvadel juhtudel põhjustavad nad pahaloormulisi anogenitaalpiirkonna kasvajaid, samuti võib larüngeaalne papillomatoos muunduda pahaloormuliseks [2].

### **2.2 Genoomne ülesehitus ja kodeeritavad valgud**

Hoolimata papilloomiviiruse erinevate tüüpide liigi- ning koespetsiifilisusest on neil kõigil sarnane genoomne ülesehitus. Võrreldes muude viirustega on papilloomiviirused suhteliselt väikesed. Nende tsirkulaarne kaheaahelaline genoomne DNA koosneb ligikaudu 8000 aluspaarist ning kodeerib kaheksat avatud lugemisraami [6].

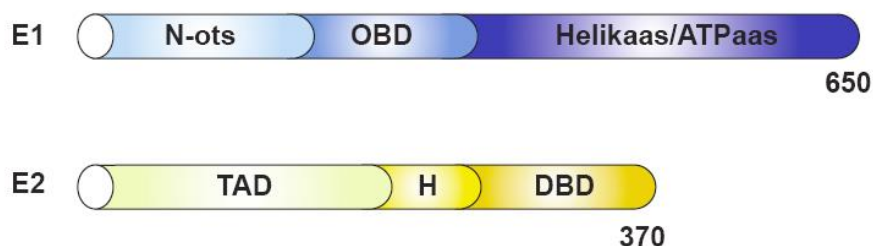
Genoom jaguneb kolmeks piirkonnaks: varajane kodeeriv (E – *early*), hiline kodeeriv (L – *late*) ja mittekodeeriv (LCR – *long control region* või URR – *upstream regulatory region*) piirkond. Genoomi piirkonnad on üksteisest eraldatud kahe polüadenülatsoonisaidiga. Varajane kodeeriv piirkond katab üle 50% genoomist ning kodeerib kuut avatud lugemisraami nn varajaste valkude (E1, E2, E4, E5, E6, E7) sünteesiks, hiline kodeeriv piirkond katab ligikaudu 40% genoomist ning kodeerib kaht avatud lugemisraami nn hiliste valkude (L1 ja L2) sünteesiks. Mittekodeeriv piirkond katab ligikaudu 10% genoomist ning sisaldab transkriptsiooniks ja replikatsiooniks vajalikke regulaatorelemente, nt replikatsiooni alguspunkti (ori – *origin of replication*), promootoreid, E1 ja E2 valgu sidumiskohti jne [6]. HPV 11 genoomi ülevaade on toodud Joonisel 1.



**Joonis 1.** HPV11 lineariseeritud genoomi ülevaade [11].

Valgud E1 ja E2 on kõige rohkem konserveerunud ja uuritud papilloomiviiruse valgud. E1 valgul on oluline roll viiruse genoomi replikatsiooniprotsessis. E1 koosneb ligikaudu 650-st aminohappe jäägist ning ta jaguneb kolmeks piirkonnaks: N-otsas asuv domään, millel puudub kindlakstehtud funktsioon, järjestusspetsiifiline DNA sidumiskompleks ja C-otsas asuv domään, mis omab helikaaset, ATPaaset ja E2 valguga seondumise aktiivsust. E1 on ainus papilloomiviiruse valk, mis omab ensümaatilist aktiivsust [7,8].

E2 koosneb ligikaudu 370-st aminohappe jäägist ning ta jaguneb samuti kolmeks piirkonnaks. N-otsas asub transaktivatsiooni domään (TAD) ning C-otsas asub DNA sidumise ja dimerisatsiooni domään. Neid kahte domääni eraldab nn hinge piirkond. E2 on multifunktsionaalne valk, mis seostub viiruse genoomi reguleeriva piirkonna spetsiifiliste sidumiskohadega ja käivitab seeläbi viiruse DNA replikatsiooni ja reguleerib transkriptsiooni. Lisaks kontrollib E2 raku jagunemisel viiruse genoomi segregatsiooni ja jaotumist tütarakkudesse [7,8]. Valkude E1 ja E2 piirkondade ülevaade on toodud Joonisel 2.



**Joonis 2.** Valkude E1 ja E2 piirkonnad [8]. Numbrid parempoolses otsas viitavad aminohappe jääkide arvule. OBD – valgu E1 DNA-ga seondumise domään, H – valgu E2 hinge piirkond.

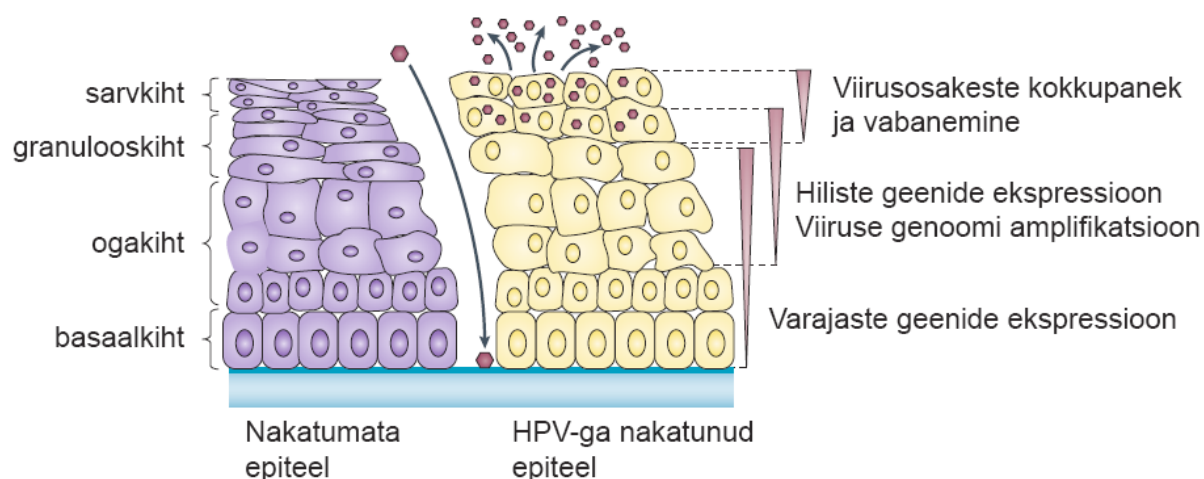
E4 ja E5 interakteeruvad peremeesraku valkudega, et muuta rakusisest keskkonda, mõjutavad rakutsükli ning diferentseerumisprogrammi. E6 ja E7 mõjutavad rakutsükli kontrollmehhanisme, et aidata viirusel vältida immuunvastust [7]. Kõrge riskiga HPV tüüpide puhul toimivad E6 ja E7 onkovalkudena. E7 seostub retinoblastoomi valkudega (kasvaja supressorvalgud) ning inhibeerib nende võimet moduleerida E2F transkriptsioonifaktorite funktsiooni. Madala riskiga HPV tüüpide puhul on näidatud, et ka nende valgud E6 ja E7 seonduvad vastavate regulaatorvalkudega, ent madalama afiinsusega [11].

Hilised valgud L1 ja L2 on kapsiidivalgud, vastavalt suur ja väike kapsiidivalg. Need valgud ekspresseeritakse viiruse elutsükli hilises staadiumis diferentseeruva epiteeli ülemistes kihtides, kus nendest moodustatakse viiruse kapsiid [6].

### 2.3 Papilloomiviiruse elutsükkel

Võrreldes teiste viirustega omab papilloomiviirus ainulaadset elutsükli. Kui enamik viiruseid paljuneb samas nakatunud rakus, siis papilloomiviiruse uute viirusosakeste süntees ja kokkupanemine toimub jagunenud nakatunud raku diferentseerivas tütararakus [10]. Kuna papilloomiviiruse genoom paljunemiseks vajalikke ensüüme ise ei kodeeri, sõltub papilloomiviiruse DNA replikatsioon peremeesraku ensüümidest. Sellest tulenevalt on papilloomiviiruse elutsükkel väga tihedalt seotud diferentseeruva epiteelraku elutsükliga.

Mitmekihiline epiteel koosneb basaalkihist ja selle kohal olevatest diferentseeruvate rakkude kihtidest. Pärast basaalaraku jagunemist kaotab üks tütarakkudest kontakti basaalmembraaniga, liigub pindmise kihi poole, väljub rakutsüklist ning lülitub diferentseerumisprotsessi. Diferentseerumisprotsessi käigus talletavad rakud keratiine, liiguvad ogakihti ja sealt edasi granulooskihti, kaotavad rakutuuma ning protsessi lõpuks sarvestuvad täielikult. Terve ja nakatunud epiteeli elutsüklike võrdlus on toodud Joonisel 3.



**Joonis 3.** Terve ja nakatunud epiteeli elutsükkel [10].

Papilloomiviiruse elutsükli saab jaotada kolmeks staadiumiks: (i) peremeesraku nakatumine ja mitteproduktiivse faasi püstitumine, (ii) mitteproduktiivse faasi säilimine ja (iii) produktiivne faas [6].

Papilloomiviirused nakatavad mitmekihilise epiteeli paljunemisvõimelisi basaalkihi rakke. Väikeste vigastuste kaudu jõuavad viirusosakesed basaalarakkudeni ning kinnituvad nende pinnal olevatele sidumiskohtadele. Kinnitumise mehhanism ei ole täpselt kindlaks tehtud, aga arvatavasti on paljude HPV tüüpide esmaseks sidumiskohaks heparaansulfaat proteoglükaan. Endotsütoosi teel siseneb viirus rakku, genoom vabaneb kapsiidist ning transporditakse rakutuuma, kus toimub viiruse DNA amplifikatsiooniline replikatsioon [9]. Esmase amplifikatsiooni tulemusena tõuseb viiruse genoomi koopiaarv 50...100 koopiani raku kohta [8]. Stabiilse replikatsiooni staadiumis püsib viiruse koopiaarv raku kohta stabiilsena, sest viiruse genoomi paljundatakse peremeesraku genoomiga peaaegu sünkroonselt. Kui nakatunud rakud jõuavad diferentseerumisprotsessi hilisesse järku ja liiguvad epiteeli pindmistesse kihtidesse, algab viiruse kontrollimatu vegetatiivne replikatsioon ehk produktiivne faas. Sünteesitakse hilised valgud ning moodustatakse viirusosakesed. Lõplikult diferentseerunud rakud surevad ning kooruvad koos viirusosakestega epiteeli pinnalt maha [6].

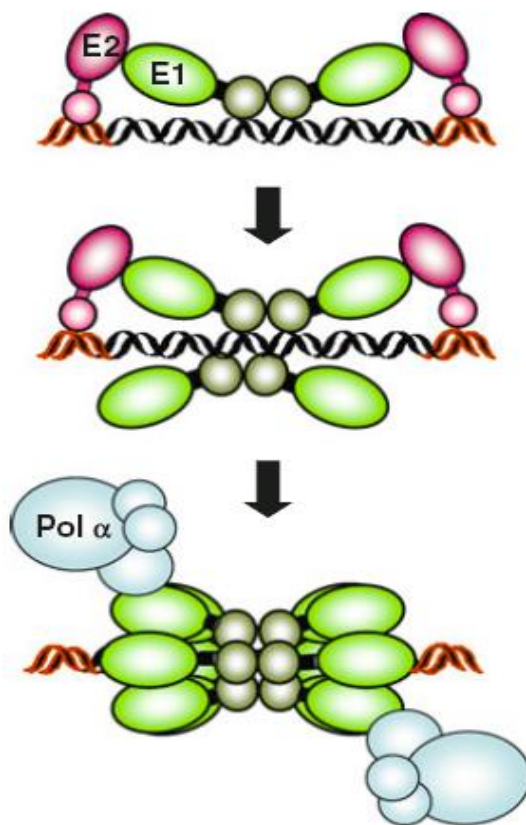
## 2.4 Papilloomiviiruse DNA replikatsioon

DNA replikatsiooni initsieerimiseks on vajalik replikatsioonikompleksi moodustumine viiruse genoomi mittekodeerivas piirkonnas asuvas replikatsiooni alguspunktis (*ori*) [6].



E1 ja E2 seonduvad omavahel E1 valgu helikaasse domääni ja E2 valgu transaktivatsiooni domääni (TAD) kaudu. Mõlemad valgud seonduvad spetsiifiliselt DNA sidumiskohadega E1BS ja E2BS. E1 dimeriseerub replikatsiooni alguspunktis ning täiendavate E1 monomeeride kaasamisega moodustub ümber DNA ahela E1 valkudest topeltheksameer [7].

Selles protsessis on ATP ja E1 vahelisel interaktsioonil kompleksne roll – ATP-ga seondumine stimuleerib E1 ja E2 kooperatiivset seostumist genoomi replikatsiooni alguspunktiga, ent muudab E1-E2 kompleksi konformatsiooni ning seeläbi nõrgendab E1 ja E2 vahelist vastasmõju. E1 heksameeri moodustumise järgselt katkeb E1 ja E2 vaheline side ning E1-ga seostub DNA polümeraas  $\alpha$ . Tekkinud kompleks asub genoomi replitseerima [7]. Papilloomiviiruse DNA replikatsiooni initsieerimine on kujutatud Joonisel 4.



**Joonis 4.** Papilloomiviiruse DNA replikatsiooni initsieerimine [7].

## 2.5 Papilloomiviiruste vastane ravi

Olemasolevate tüügaste puhul kasutatakse tavaliselt paikset medikamentooset või kirurgilist ravi. Kondüloomide paikne medikamentoosne ravi hõlmab tsütotoksilise aine

podofüllotoksiin või TLR-7 retseptori (*toll-like receptor 7*) agonisti kasutamist. Soolatüügaste puhul kasutatakse ka salitsüül- ja piimahappe lahust. Kirurgiline ravi hõlmab krüoteraapiat, elektrokoagulatsiooni, silmuskonisatsiooni, süsinikdioksiidlaseriga või kirurgilist eemaldamist [2].

Eelpool toodud ravimeetodid on paraku puudulikud. Peamiseks puuduseks on ravi madal efektiivsus – väga paljudel juhtudel võivad tüükad mõne kuu jooksul pärast eemaldamist taastekkida, eriti immuunpuudulikkusega patsientidel. Samuti on paikne ravi seotud sagedasti valuga või kõrvaltoimetega nagu lööve ja põletik [7].

## **2.6 HPV valgud ravimite sihtmärkidena**

HPV-del puuduvad ensüümid, mis on tavaliselt viirusevastaste ainete sihtmärgiks (nt proteaasid ja polümeraasid) [8]. See on seadnud HPV-vastaste ravimite väljatöötamisele teatavad piirangud, kuid täiustunud teadmised papilloomiviiruse elutsükli kohta on andnud võimaluse uurida papilloomiviiruse valke potentsiaalsete HPV-vastaste ravimite sihtmärkidena.

Valkude E4 ja E5 funktsiooni kohta on hetkel liiga vähe teavet, et kasutada neid sobivate kandidaatidena HPV-vastaste ravimite uurimisel. Valkude E6 ja E7 funktsiooni ja toimemehhanismide kohta peremeesrakkudes on praeguseks rohkem teavet, mistõttu on need valgud HPV-vastaste ainete otsimisel sobivamad sihtmärgid. Valkude E6 ja E7 puuduseks on see, et nad ei ole kõrgelt konserveerunud. Sel põhjusel ei pruugi mingi kindla HPV tüübi E6 või E7 valgu inhibiitor omada ravitoimet kõikide oluliste HPV tüüpide vastu [7].

E1 on ainus papilloomiviiruse valk, mis omab ensümaatilist aktiivsust (ATPaasne ja DNA helikaasne aktiivsus). Samuti on E1 kõige rohkem konserveerunud papilloomiviiruse valk [3]. Papilloomiviiruse DNA replikatsiooni mehhanismi juures toodud valkudega E1 ja E2 seotud sammud on potentsiaalsed sihtmärgid inhibeerimisele [7].

Praeguseks ajaks on leitud suhteliselt vähe madalamolekulaarseid E1–E2 valkude vahelise interaktsiooni inhibiitoreid. Oletatavasti seetõttu, et madalamolekulaarsetel ühenditel on raske konkureerida suure pinnaga valkude omavahelise interaktsiooniga ning valkude pinnad on suhteliselt siledad ehk seal puuduvad väikestele molekulidele sobivad sidumistaskud [8].

Seetõttu on käesolevas töös modelleeritud E1 helikaasi ATPaasse aktiivsuse inhibeerimist madala molekulmassiga ühenditega.

Esitatud mudel peaks leidma rakendust sarnase struktuuriga ühendite inhibeerimisvõime arvutuslikul hindamisel, et lihtsustada ja toetada uute võimalike inhibiitorite sünteesimiseks ja farmakoloogiliste parameetrite eksperimentaalseks mõõtmiseks tehtavat tööd.

### 3. Arvutuslik osa

Antud töö arvutuslikus osas kasutati kvantitatiivse struktuur-aktiivsuse sõltuvuse (QSAR) meetodit, et arendada olemasolevate eksperimentaalsete andmete põhjal uuritavat omadust kirjeldav mudel.

#### 3.1 Kvantitatiivse struktuur-aktiivsuse sõltuvuse (QSAR) meetod

QSAR meetodi aluseks on seos ühendi molekulaarse struktuuri ja aktiivsuse/omaduste vahel. Matemaatilise valemiga väljendatuna:

$$\text{Aktiivsuse/omadus} = f(D),$$

kus D tähistab molekule kirjeldavate parameetrite, ehk deskriptorite komplekti.

QSAR mudeli rakendatavus sõltub eksperimentaalsete andmete täpsusest, deskriptorite ja statistiliste meetodite valikust ning valideerimisest. Kasulik mudel on hästi interpreteeritav ja sisaldab optimaalsel hulgal relevantseid deskriptoreid [13].

QSAR mudeli loomise võib jaotada kolmeks suuremaks etapiks: (i) andmete ettevalmistamine ja analüüs, (ii) deskriptorite arvutamine, (iii) mudelite arendamine ja valideerimine [13].

#### 3.2 Molekulaardeskriptorid

Formaalselt on molekulaardeskriptor molekuli matemaatiline esitus, mis on saadud kindlaksmääratud molekulaarsele struktuurile kindla algoritmi rakendamise või eksperimentaalse protseduuri teel. Sellest definitsioonist tulenevalt võivad deskriptorid olla empiirilised või teoreetilised. Empiirilised deskriptorid on tavaliselt ühendi mingi füüsiko-keemilise omaduse eksperimentaalselt mõõdetud väärtused või viimaste järgi arvutatud väärtused. Teoreetilised deskriptorid arvutatakse molekuli valemist kindlate algoritmide järgi [13]. Klassikaliselt jagatakse teoreetilised deskriptorid tuletusmeetodi järgi järgmistesse rühmadesse: konstitutsionaalsed, topoloogilised, geomeetrilised, elektrostaatiliselt ja kvantkeemilised.

Konstitutsionaalsed deskriptorid on kõige lihtsamad molekule kirjeldavad deskriptorid. Need tuletatakse otseselt ühendi valemist ning need annavad informatsiooni

molekuli keemilise koostise kohta, ent ei anna teavet molekuli geomeetrilise struktuuri kohta. Konstitutsioonilised deskriptorid on näiteks aatomite koguarv, mingi elemendi aatomite või funktsionaalrühmade kogu- või suhtarv, sidemete või tsüklite kogu- või suhtarv, molekulmass jne [12, 13].

Topoloogilised deskriptorid kirjeldavad aatomitevahelisi sidemeid ja molekulide hargnemist. Need tuletatakse molekuli kahemõõtmelisest struktuurist graafiteooria alusel. Tuntuimad topoloogilised deskriptorid on Balaban'i indeks, Kier'i ja Hall'i indeksid, Wiener'i indeks, Randić'i indeksid, informatsioonisisalduse indeksid jne [12].

Geomeetrilised deskriptorid kirjeldavad atomite kolmemõõtmelist paiknemist molekulis. Need tuletatakse aatomite kolmemõõtmeliste koordinaatide põhjal. Olulisemad geomeetrilised deskriptorid on molekuli pindala, maht, inertsimoment, gravitatsiooniindeksid jne [12].

Elektrostaatiliselt kirjeldavad molekuli laengujaotust. Need tuletatakse poolempiiriliste või empiiriliste meetodite abil arvutatud osalaengute põhjal. Elektrostaatiliste deskriptorite hulka kuuluvad osalaengud, polaarsusparameetrid, polariseeritavus, elektrostaatiline potentsiaal molekuli pinnal, dipoolmoment, ionisatsioonenergia jne. Elektrostaatiliste deskriptorite alla kuuluvad ka laetud osapindalaga seotud deskriptorid (CPSA). CPSA deskriptorid tuletatakse molekulide pindala ja aatomite osalaengute põhjal ning neid kasutatakse molekulidevaheliste polaarse iseloomuga vastasmõjude kirjeldamiseks [12,14].

Kvantkeemilised deskriptorid kirjeldavad molekuli elektroonseid ja geomeetrilisi omadusi ning molekulisiseseid vastasmõjusid. Need tuletatakse molekuli elektronide lainefunktsiooni kvantkeemilistest arvutustest. Olulisemad kvantkeemilised deskriptorid on molekuli koguenergia, protoneerumise energia, osakeste omavaheliste interaktsioonide energiad jne. Kvantkeemilised deskriptorite alarühmadeks on molekulaarorbitaalidega seotud deskriptorid (nt HOMO ja LUMO energiad, reaktiivsused, sideme järgud, vaba valents jne.) ja termodünaamilised deskriptorid (nt molekuli vibratsiooniline ja translatsiooniline entalpia, vibratsiooniline, rotatsiooniline ja translatsiooniline entroopia ja vibratsiooniline soojusmahtuvus) [12].

Ülaltoodud jaotus on tinglik ning deskriptoreid võib jaotada nende olemuse järgi ka muude printsiipide alustel. Kuna antud töös kasutati deskriptorite arvutamisel CODESSA PRO tarkvara, on deskriptorid jaotatud ülaltoodud rühmadesse CODESSA PRO-s kasutatava jaotuse alusel [12].

## 4. Andmekomplekt ja metoodika

### 4.1 Andmekomplekt

Antud töös kasutatud eksperimentaalsed andmed 42 ühendi jaoks pärinevad Faucheri ja tema kaastöötajate poolt 2004. a avaldatud artiklist [3]. Faucher jt töö eesmärgiks oli eksperimentaalse struktuur-aktiivsus sõltuvuse meetodiga leida (bifenüül-4-sulfonüül)äädikhappe (vt. Lisa 1, ühend 18) derivaatide hulgast E1 valgu ATPaase aktiivsuse inhibiitorit. Uuritav omadus, mida ka antud töös eksperimentaalselt mõõdetud omaduse ehk aktiivsusega kasutatakse, on  $IC_{50}$  – kontsentratsioon, mille juures saavutatakse pool maksimaalsest inhibeerimisest.

Eksperimentaalsed andmed on teisendatud logaritmilisele skaalale, võttes molaarsest kontsentratsioonist negatiivse logaritmi ( $pIC_{50} = -\log (IC_{50})$ ). Seega tuleb tulemuste interpreteerimisel silmas pidada, et suurem  $pIC_{50}$  väärtus viitab madalamale inhibeerimise saavutamiseks vajalikule kontsentratsioonile ehk kõrgemale aktiivsusele. Ühendite struktuurivalemid, mõõdetud ja ennustatud  $pIC_{50}$  väärtused ning mõõdetud ja ennustatud väärtuste vahe on toodud Lisas 1.

### 4.2 Metoodika

Ühendite struktuurifailid ja omadused võeti ChEMBL andmebaasist [15]. 3D struktuurid eeloptimeeriti kasutades MM+ [16] molekulaarmehaanilist jõuvälja. Programmi OpenBabel [17] kasutati madalaima energiaga konformeerimise leidmiseks. Lõplik optimeerimine teostati CODESSA PRO-sse implementeeritud MOPAC 7.05 (kloon) [18] abil, kasutades poolempiirilist meetodit AM1 (*Austin Model I*) [19].

Seejärel arvutati CODESSA PRO abil konstitutsionaalsed, topoloogilised, geomeetrilised, elektrostaatilised ja kvantkeemilised molekulaardeskriptorid, kokku 836 deskriptorit.

Arvutatud deskriptorite põhjal arendati regressioonimudelid, kasutades parima multilineaarse regressiooni (BMLR – *best multilinear regression*) algoritmi. BMLR meetod koosneb alljärgnevatest sammudest.

- (1) Andmehulgast otsitakse kõik ortogonaalsete deskriptorite paarid;
- (2) Ortogonaalsete deskriptorite paaride põhjal luuakse kahedeskriptorilised võrrandid ning nende hulgast valitakse kõrgeima korrelatsioonikoefitsiendiga ( $R^2$ ) võrrandid.

Edasiseks arendamiseks kasutatakse neljasadat parima korrelatsioonikoefitsiendiga võrrandit;

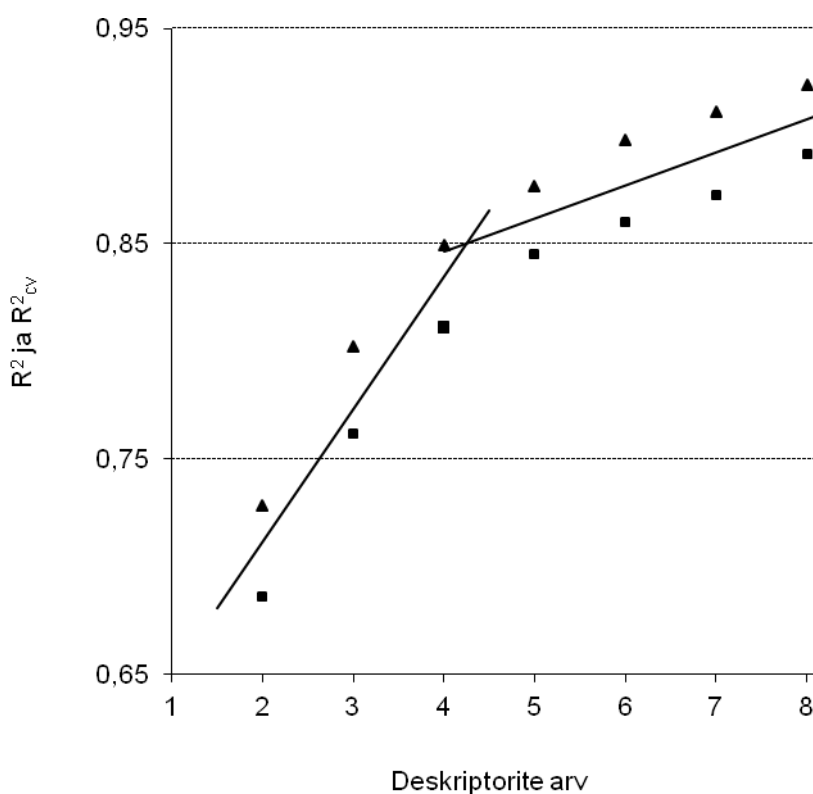
(3) Eelnevalt leitud kahedeskriptorilisele võrrandile lisatakse uusi deskriptoreid seni, kuni Fisheri kriteerium ( $F$ ) muutub väiksemaks kui parima korrelatsioonikoefitsiendiga kahedeskriptorilise võrrandi oma. Protsessi korratakse üle kõikide paaride ning parima korrelatsioonikoefitsiendiga ( $R^2$ ) võrrand salvestatakse;

(4) Eelmises punktis leitud kolmedeskriptorilisele võrrandile lisatakse uusi deskriptoreid seni kuni lisamine enam ei paranda Fisheri kriteeriumi ning parima korrelatsioonikoefitsiendi ( $R^2$ ), ristvalideeritud korrelatsioonikoefitsiendi ( $R^2_{cv}$ ) ja Fisheri kriteeriumiga ( $F$ ) võrrand salvestatakse.

Deskriptorite paare, mis olid omavahel korreleerunud ( $R^2 > 0,6$ ), peeti mudeli arendamiseks ebasobivaks ning üks deskriptoritest eemaldati. Samuti eemaldatakse deskriptorid, millele tulenevalt ühendite iseärasustest ei ole väärtust arvatud (nt deskriptor, mille arvutamisel kasutatav jagaja võrdub nulliga).

## 5. Tulemused ja analüüs

Eelmises punktis kirjeldatud metoodika alusel arendati mudelid kahest kuni kaheksa deskriptorini. Optimaalse deskriptorite arvuga mudeli leidmiseks analüüsiti deskriptorite lisamise panust korrelatsioonikoefitsientide paranemisse. Joonisel 5 on toodud deskriptorite arv ja vastavate mudelite korrelatsioonikoefitsiendid ning ristvalideeritud korrelatsioonikoefitsiendid. Jooniselt ilmneb, et murdepunkt asub neljanda deskriptori juures, mis näitab, et neljandast deskriptorist alates ei too deskriptorite arvu suurenemine kaasa mudeli märkimisväärt statistilist paranemist.



**Joonis 5.** Deskriptorite arv ja nendele vastavate mudelite korrelatsioonikoefitsiendid.

Kolmnurkadega on tähistatud korrelatsioonikoefitsiendid ( $R^2$ ), ruutudega ristvalideeritud korrelatsioonikoefitsiendid ( $R^2_{cv}$ ).

Sobivaima mudeli välja selgitamiseks jätkati analüüsi kolme kuni viit parameetrit sisaldavate mudelitega.



Kolmeparameetrilise mudeli korrelatsioonikoefitsient on tunduvalt väiksema väärtusega kui nelja deskriptoriga mudeli korral, mis viitab mudeli madalamale ennustamisvõimele. Lisaks esines kolme deskriptoriga mudeli korral üks hälbiv punkt (ühend 04, vt Lisa 1).

Neljanda deskriptori lisamisel paranesid mudeli statistilised näitajad tunduvalt. Neljaparameetrilise mudeli korral omasid deskriptorid sisukaid väärtusi ning jaotusid hästi. Deskriptorite väärtuste jaotumine on esitatud Lisas 3. Neljaparameetrilises mudelis hälbivaid punkte ei olnud.

Samas viie parameetriga mudeli puhul ei täheldatud mudeli olulist statistilist paranemist võrreldes neljaparameetrilise mudeliga. Sellises olukorras on ka oht mudeli üleparameetriseerimiseks, ehk lisatanduvad deskriptorid küll parandavad korrelatsiooni, kuid ei oma seost uuritava omadusega. Nelja- ja viieparameetriliste mudelite deskriptorite võrdlusel selgus, et lisandunud deskriptori ehk viieparameetrilise mudeli Studenti t-testi järgi tähtsusest 5. deskriptori „hapniku aatomite maksimaalne osalaeng” (*Max partial charge (Zefirov) for atoms for atom O*) väärtused omasid identseid väärtuseid enamiku ühendite jaoks, mistõttu antud deskriptori kaasamine oleks muutnud mudeli uuritava omaduse kirjeldamisel vähem usaldusväärseks. Hapniku aatomite maksimaalse osalaengu väärtuste jaotumine on esitatud Lisas 4. Viieparameetrilises mudelis esines samuti üks hälbiv punkt (ühend 13, vt Lisa 1).

Ülaltoodut arvesse võttes, peeti nelja deskriptoriga mudelit kõige sobivamaks, et kirjeldada uuritavat omadust. Tabelis 1 on toodud vastav mudel koos statistiliste näitajatega, kolme- ja viieparameetrilised mudelid on toodud Lisas 2. Deskriptorid on järjestatud Studenti t-testi absoluutväärtuse alusel, mis väljendab deskriptorite olulisust mudelis [20]. Mõõdetud ja neljaparameetrilise mudeli põhjal ennustatud  $pIC_{50}$  väärtuste omavaheline korrelatsioon on toodud Joonisel 6.

**Tabel 1.** Neljaparameetrilise mudeli liikmed ning statistilised näitajad.

#	Koefitsient	t-test	Deskriptor
0	5,390	4,50	Vabaliige
1	0,219	14,13	Bonding Information content (order 2)
2	-0,040	-5,99	HA dependent HDSA-1
3	0,026	4,26	HDSA H-donors surface area
4	-10,843	-4,25	FPSA-1 Fractional PPSA (PPSA-1/TMSA)

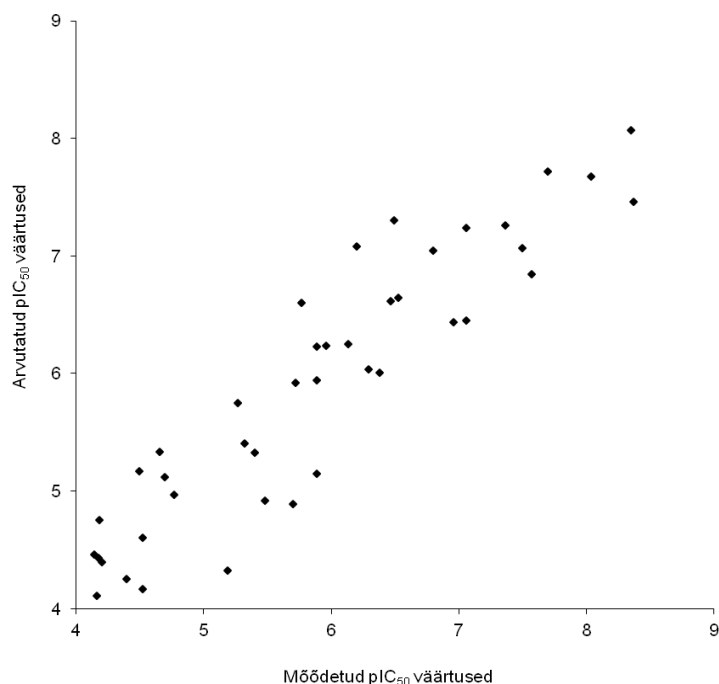
$N = 42$ ;  $n = 4$ ;  $R^2 = 0,849$ ;  $R^2_{cv} = 0,811$ ;  $F = 52,21$ ;  $s^2 = 0,25$ .

$N$  – ühendite arv,  $n$  – deskriptorite arv,  $R^2$  – korrelatsioonikoefitsient,  $R^2_{cv}$  – ristvalideeritud korrelatsioonikoefitsient,  $F$  – Fisheri kriteerium,  $s^2$  – standardhälbe ruut.

Studenti t-testi absoluutväärtuse järgi on kõige olulisem deskriptor topoloogiliste deskriptorite hulka kuuluv 2. järku sidemeinformatsioonisalduse indeks (*Bonding Information Content (order 2)*). See deskriptor kirjeldab molekulide suurust ning hargnevust [21]. Positiivne märk muutuja ees viitab sellele, et ühendi sümmeetria kahanedes inhibeeriv omadus suureneb.

Ülejäänud kolm deskriptorit kuuluvad kõik laetud osapindalaga seotud (*CPSA*) elektrostaatiliste deskriptorite hulka. Vesiniksideme doonori osapindalasid kirjeldavad deskriptorid (*HDSA*) ja (*HDSA-1*) on seotud ühendite vesiniksideme moodustamise võime ja molekulidevahelise vastasmõjuga. *HDSA* on defineeritud kui vesiniksideme doonoriks olevate vesinikuaatomite kogupindala. *HDSA-1* vesiniksideme aktseptorist sõltuv vesiniksideme doonori pindala [12, 22]. Suurema vesiniksideme doonori pindala korral on vesiniksideme moodustumine lihtsam. Nende deskriptorite esinemine mudelis võib olla seletatav sulfonüülgrupi esinemisega uuritavates ühendites.

Neljas deskriptor, positiivselt laetud osapindala (*FPSA-1*), on laenguga seotud deskriptor, mis on määratletud kui molekulide positiivse osapindala ja kogupindala suhe [12]. See deskriptor näitab, kuidas laeng on molekulide kogupindala suhtes jaotatud ja on seostatav üldise reaktsioonivõimega, ning sarnaselt kahele eelmisele deskriptorile võib ka seda seostada vesiniksideme moodustumise võimega.



**Joonis 6.** Mõõdetud ja arvutatud  $pIC_{50}$  väärtused.

Deskriptorite omavahelise korrelatsiooni puudumise välja selgitamiseks uuriti deskriptorite korrelatsioonikoefitsiente ( $R^2$ ). Neljaparaameetrilise mudeli deskriptorite omavahelise sõltuvuse korrelatsioonikoefitsiendid on toodud Tabelis 2.

**Tabel 2.** Neljaparaameetrilise mudeli deskriptorite interkorrelatsioon ( $R^2$ ).

	<b><sup>2</sup>BIC</b>	<b>HDSA-1</b>	<b>HDSA</b>	<b>FPSA-1</b>
<b><sup>2</sup>BIC</b>	1	0,29	0,02	0,28
<b>HDSA-1</b>	0,29	1	0,41	0,18
<b>HDSA</b>	0,02	0,41	1	0,07
<b>FPSA-1</b>	0,28	0,18	0,07	1

Tabelis olevad väärtused kinnitavad, et deskriptorite omavaheline korrelatsioon on tugevalt väiksem kui määratud kriteerium ( $R^2 < 0,6$ ). Suurim korrelatsioon esineb deskriptorite HDSA ja HDSA-1 vahel. Mõlemad deskriptorid kirjeldavad üldiselt sama omadust – vesiniksideme moodustamise võimet ja peamiseks erinevuseks on, et HDSA-1 arvestab ka vesiniksideme aktseptori mõju. Ülejäänud deskriptoritevahelised korrelatsioonikoefitsiendid on tunduvalt väiksemad. Selle põhjal saab olla kindel, et mudelis olevad deskriptorid on ortogonaalsed ning mudel on statistiliselt usaldusväärne.

## 5.1 Mudeli valideerimine

Juhusliku korrelatsiooni esinemise väljaselgitamiseks valideeriti mudel sisemiselt nn ABC valideerimismeetodiga [23, 24]. Mudeli loomisel kasutatud ühendid jaotati kolme rühma: ühendid 1, 4, 7 jne moodustasid rühma A, ühendid 2, 5, 8 rühma B ning ühendid 3, 6, 9 rühma C. Ühendid jaotati tekkinud rühmade A, B ja C alusel treening- ja testrühmadeks. Näiteks rühma A ja B ühendeid kasutati mudeli treenimiseks ning rühma C ühendeid mudeli testimiseks. Analoogselt testiti rühmi A ja B, kasutades rühmi B + C ja A + C vastavate treeningrühmadena. Kuna ühendid on järjestatud  $pIC_{50}$  kasvamise järjekorras, siis tagab selline jaotamine ühendite ühtlase jaotumise treening- ja arvutatavate rühmade vahel. Arendati parima multilineaarse regressiooni mudelid kasutades neljaparaameetrilises mudelis olevaid deskriptoreid, lastes muutuda ainult koefitsientidel. Saadud mudelite korrelatsioonikoefitsiendid on toodud Tabelis 3.

**Tabel 3.** ABC valideerimise treening- ja arvutatavad rühmad ning statistilised parameetrid.

Treeningrühm	$R^2$	$s^2$	Arvutatav rühm	$R^2$	$s^2$
B + C	0,865	0,25	A	0,842	0,34
A + C	0,864	0,24	B	0,905	0,22
A + B	0,844	0,28	C	0,896	0,25
A + B + C	0,849	0,25			

$R^2$  – korrelatsioonikoefitsient,  $s^2$  – standardhälbe ruut.

Tabelis 3 toodud andmed viitavad heale korrelatsioonile. Seega lubavad tabelis toodud statistilised näitajad eeldada, et antud mudel on usaldusväärne ja hea ennustusvõimega.

## 5.2 Mudeli kasutusala

Mudeli loomiseks kasutatud (bifenüül-4-sulfonüül)äädikhappe derivaatide eksperimentaalsed  $pIC_{50}$  väärtused jäävad vahemikku 4,14 – 8,37. Struktuuralse varieerumise ning mõõdetud ja ennustatud  $pIC_{50}$  piirväärtuste suurenemisel ei pruugi antud mudel nende inhibeerivaid omadusi relevantset ennustada.

## 6. Kokkuvõte

Käesolevas töös on antud ülevaade inimese papilloomiviiruse olemusest ja HPV-nakkusega seotud haiguste ravivõimalusest. Rohkem rõhku on asetatud nende elutsüklile, DNA replikatsioonile ja replikatsiooniprotsessis osalevatele valkudele.

Töö arvutuslikus osas loodi (bifenüül-4-sulfonüül)äädikhappe derivaatidel põhinev inimese papilloomiviiruse (HPV6) E1 helikaasi ATPaasse aktiivsuse inhibeerimist kirjeldav QSAR mudel. Andmekomplekt koos 42 ühendi mõõdetud  $IC_{50}$  väärtustega pärineb Faucheri ja tema kaastöötajate 2004. a avaldatud artiklist.

Sisemine valideerimine kinnitas, et leitud nelja deskriptoriga mudel kirjeldab uuritavat omadust heade statistiliste näitajatega. Regressiooni korrelatsioonikoefitsient ( $R^2$ ) on 0,849, ning ristvalideeritud korrelatsioonikoefitsient ( $R^2_{cv}$ ) 0,811. Andmekomplekt allus analüüsile hästi – leitud neljaparameetrilises mudelis hälbijaid ei olnud.

Mudelis esinevad deskriptorid on uuritava omaduse kvantiseerimisel relavantsed, kirjeldades molekuli kuju ja suurust, vesiniksideme moodustumise võimet ning laengujaotust. Kõiki neid omadusi võib seostada ligandi (madala molekulmassiga ühendid) ja ensüümi (ATPaasse aktiivsusega valk E1) vahelise vastastoimega.

Ülaltoodu põhjal saab öelda, et leitud neljaparameetriline mudel kirjeldab uuritavat omadust usaldusväärselt ning on hea ennustusvõimega treeningrühmas olevatele sarnastele ühenditele.

## 7. Summary

### QSAR Modeling of ATPase Inhibition Activity of Human Papillomavirus HPV6 E1 Helicase

**Hardi Haljas**

Current thesis provides an overview of human papillomaviruses and treatment possibilities of HPV infection related diseases. HPV life cycle, DNA replication and HPV proteins taking part in replication process are emphasised.

QSAR model describing inhibition of ATPase activity of human papillomavirus (HPV6) E1 helicase with (biphenyl-4-sulphonyl)acetic acid derivatives was developed in computational section. Data set containing experimentally measured  $IC_{50}$  values for 42 compounds were taken from work of Faucher *et al* published in 2004.

Internal validation confirmed that developed four-parameter model describes the studied property with good statistical characteristics. Correlation coefficient ( $R^2$ ) of regression is 0,849, cross-validated correlation coefficient ( $R^2_{cv}$ ) is 0,811. Data set subjected well to the analysis. There were no outliers in developed four-parameter model.

Descriptors represented in model are relevant for determining the studied property as these describe shape and size of molecules, hydrogen bond formation capability and charge distribution. All these characteristics can be associated with interaction between ligand (small-molecule compounds) and enzyme (E1 protein possessing ATPase activity).

Based on above, we may propose that developed four-parameter model describes studied property reliably and has good predictive power for similar compounds to the training set.

## 8. Kasutatud kirjandus

1. The Community Register of Medicinal Products of European Commission, [http://ec.europa.eu/health/documents/community-register/index\\_en.htm](http://ec.europa.eu/health/documents/community-register/index_en.htm) (viimati alla laetud 24.05.2013).
2. Lacey, C. J. N.; Lowndes, C. M.; Shah, K. V. Chapter 4: Burden and Management of Non-cancerous HPV-related Conditions: HPV-6/11 disease. *Vaccine* **2006**, *24*, Supplement 3: S35-41.
3. Faucher, A. M.; White, P. W.; Brochu, C.; Grand-Maître, C.; Rancourt, J.; Fazal, G. Discovery of Small-Molecule Inhibitors of the ATPase Activity of Human Papillomavirus E1 Helicase. *J. Med. Chem.* **2004**, *47*, 18-21.
4. Virus Taxonomy of International Committee on Taxonomy of Viruses: 2012 Release, <http://ictvonline.org/virusTaxonomy.asp?version=2012> (viimati alla laetud 22.04.2013).
5. PaVE: Papilloma Virus Genome Database, <http://pave.niaid.nih.gov> (viimati alla laetud 22.04.2013).
6. *Encyclopedia of Virology*, 3rd ed.; Mahy, B.W.J.; Regenmortel, M.H.V., Ed; Elsevier, 2008; Vol. 4, pp 8-26, 34-39.
7. White, P. W.; Faucher, A. M.; Goudreau, N. Small Molecule Inhibitors of the Human Papillomavirus E1–E2 Interaction. In *Small-Molecule Inhibitors of Protein-Protein Interactions*; Vassilev L.; Fry D., Ed; Springer – Verlag, 2011; pp 61-65.
8. D'Abramo, C. M.; Archambault, J. Small Molecule Inhibitors of Human Papillomavirus Protein - Protein Interactions. *Open Virol. J.* **2011**, *5*, 80-95.
9. Doorbar, J. The Papillomavirus Life Cycle. *J. Clin. Virol.* **2005**, *32*, Supplement 1: S7-15.
10. Moody, C. A.; Laimins, L. A. Human Papillomavirus Oncoproteins: Pathways to Transformation. *Nat. Rev. Cancer* **2010**, *10*, 550-560.
11. Oh, S. T.; Longworth, M. S.; Laimins, L. A. Roles of the E6 and E7 Proteins in the Life Cycle of Low-Risk Human Papillomavirus Type 11. *J. Virol.* **2004**, *78*, 2620-2626.
12. CODESSA PRO, <http://www.codessa-pro.com>
13. Puzyn, T.; Leszczynski, J.; Cronin, M. T. D., Ed.; *Recent Advances in QSAR Studies: Methods and Applications*; Springer Science+Business Media B.V., 2010; pp 5-9, 29-30.
14. Stanton, D. T.; Jurs, P. C. Computer-Assisted Study of the Relationship between Molecular Structure and Surface Tension of Organic Compounds. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 109-115.

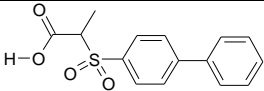
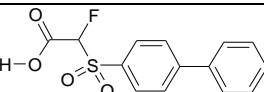
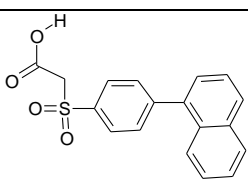
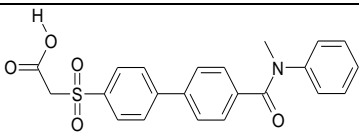
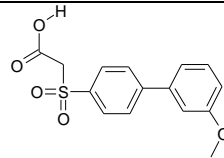
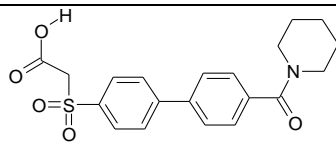
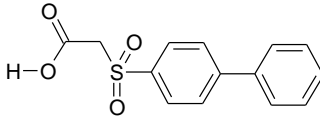
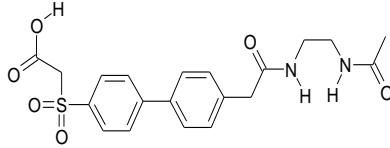
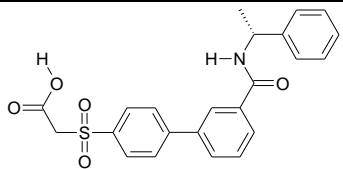
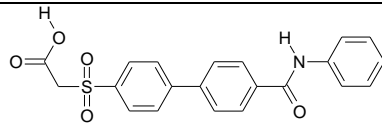
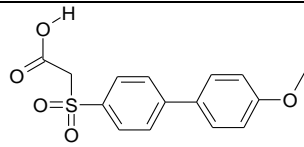
15. Chemogenomics Database of European Molecular Biology Laboratory; <https://www.ebi.ac.uk/chembl/> (viimati alla laetud 06.05.2013).
16. Hocquet, A.; Langg rd, M. An Evaluation of the MM+ Force Field. *J. Mol. Model.* **1998**, *4*, 94-112.
17. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminformatics* **2011**, *3*, 14.
18. Stewart, J. J. P. MOPAC: A Semiempirical Molecular Orbital Program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1-103.
19. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.
20. StatSoft Electronic Statistics Textbook, <http://www.statsoft.com/textbook/>
21. Basak, S. C.; Harriss, D. K.; Magnuson, V. R. Comparative Study of Lipophilicity versus Topological Molecular Descriptors in Biological Correlations. *J. Pharm. Sci.* **1984**, *73*, 429-437.
22. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, 2000; pp 53-56.
23. T mm, K.; Fara, D. C.; Katritzky, A. R.; Burk, P.; Karelson, M. A Quantitative Structure–Property Relationship Study of Lithium Cation Basicities. *J. Phys. Chem. A* **2004**, *108*, 4812-4818.
24. Katritzky, A. R.; Tatham, D. B.; Maran, U. Theoretical Descriptors for the Correlation of Aquatic Toxicity of Environmental Pollutants by Quantitative Structure-Toxicity Relationships. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1162-1176.

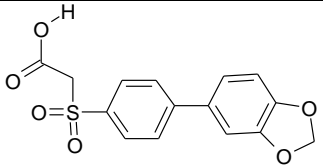
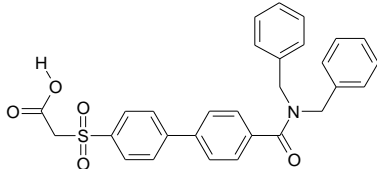
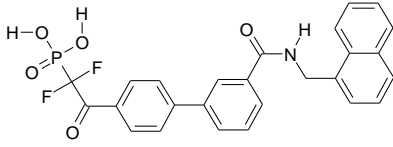
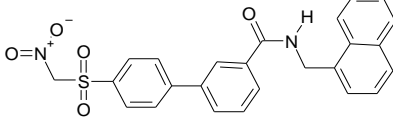
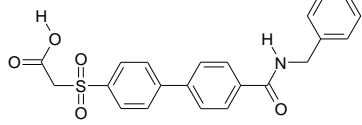
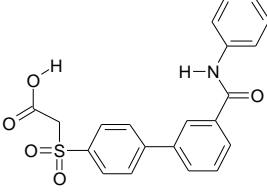
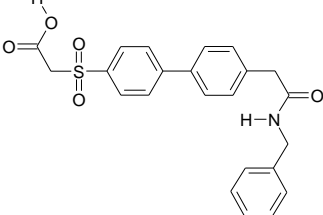
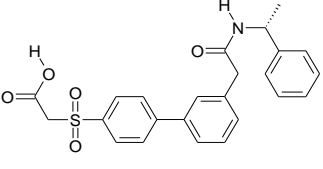
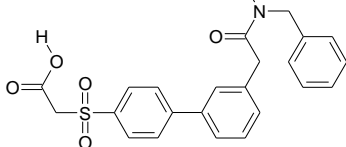


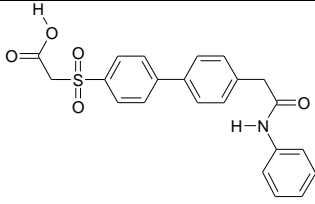
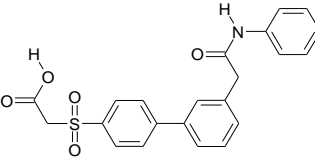
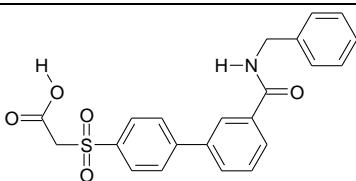
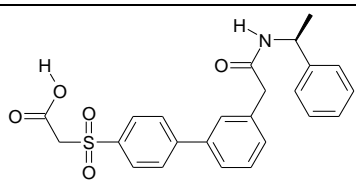
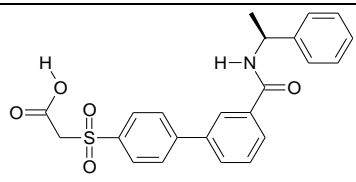
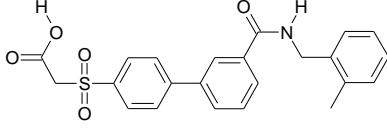
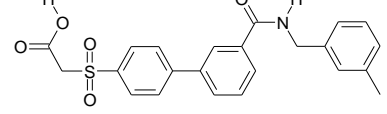
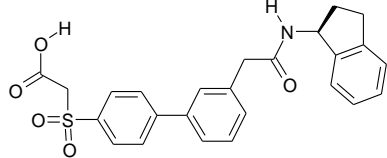
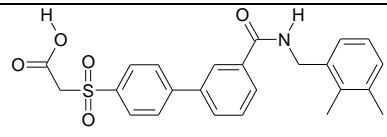
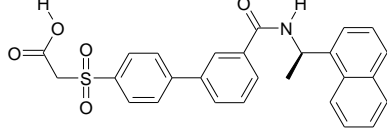
## Lisad

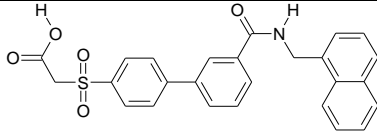
**Lisa 1.** Ühendite struktuurivalemid ning eksperimentaalne ja ennustatud IC<sub>50</sub>.

Ühendi number	Ühendi struktuurivalem	Mõõdetud pIC <sub>50</sub> , M	Ennustatud pIC <sub>50</sub> , M	Jääk-liige
01		4,14	4,46	-0,32
02		4,16	4,11	0,05
03		4,17	4,43	-0,25
04		4,19	4,75	-0,57
05		4,20	4,39	-0,19
06		4,40	4,25	0,15
07		4,49	5,17	-0,67
08		4,52	4,16	0,36
09		4,52	4,60	-0,08
10		4,66	5,34	-0,68
11		4,70	5,12	-0,42

12		4,77	4,97	-0,20
13		5,19	4,33	0,86
14		5,27	5,75	-0,48
15		5,32	5,40	-0,08
16		5,40	5,32	0,07
17		5,48	4,92	0,56
18		5,70	4,89	0,81
19		5,72	5,92	-0,20
20		5,77	6,60	-0,83
21		5,89	6,23	-0,34
22		5,89	5,14	0,74

23		5,89	5,94	-0,05
24		5,96	6,24	-0,28
25		6,13	6,25	-0,12
26		6,20	7,08	-0,88
27		6,29	6,03	0,26
28		6,38	6,01	0,37
29		6,47	6,61	-0,14
30		6,49	7,30	-0,81
31		6,52	6,64	-0,12

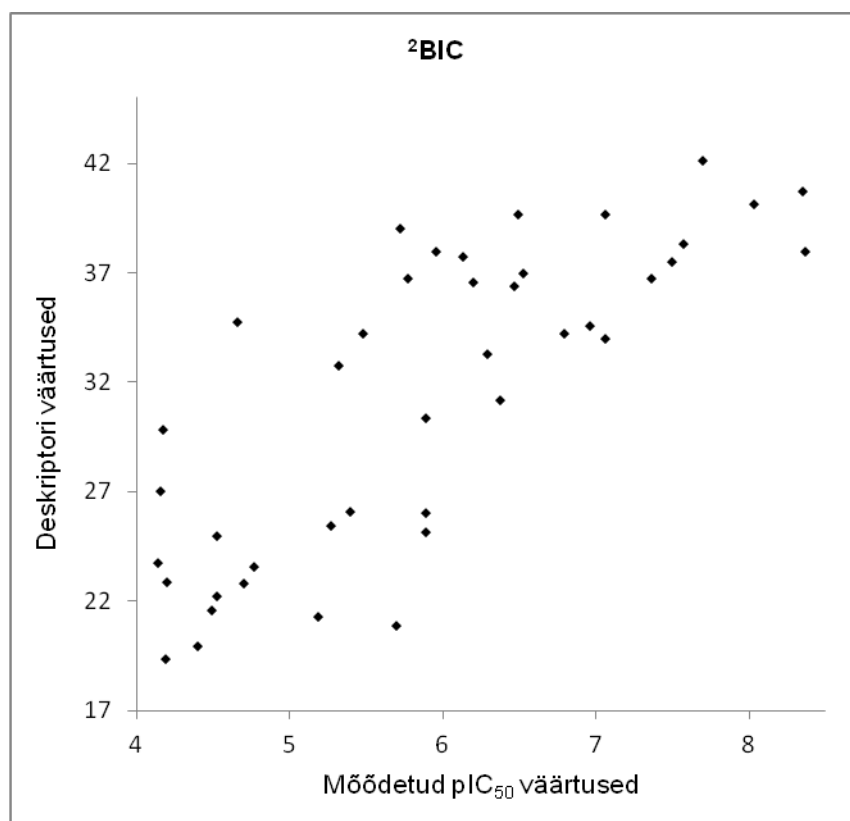
32		6,80	7,05	-0,25
33		6,96	6,44	0,52
34		7,06	6,45	0,61
35		7,06	7,24	-0,18
36		7,37	7,26	0,10
37		7,49	7,06	0,43
38		7,57	6,84	0,73
39		7,70	7,72	-0,02
40		8,04	7,68	0,36
41		8,35	8,07	0,27

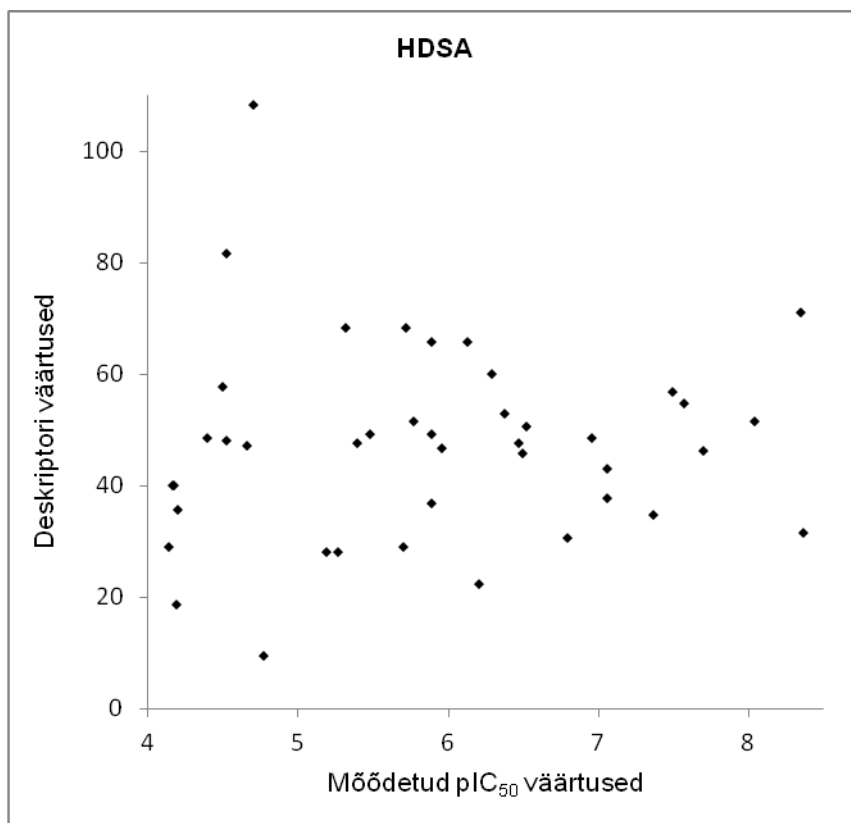
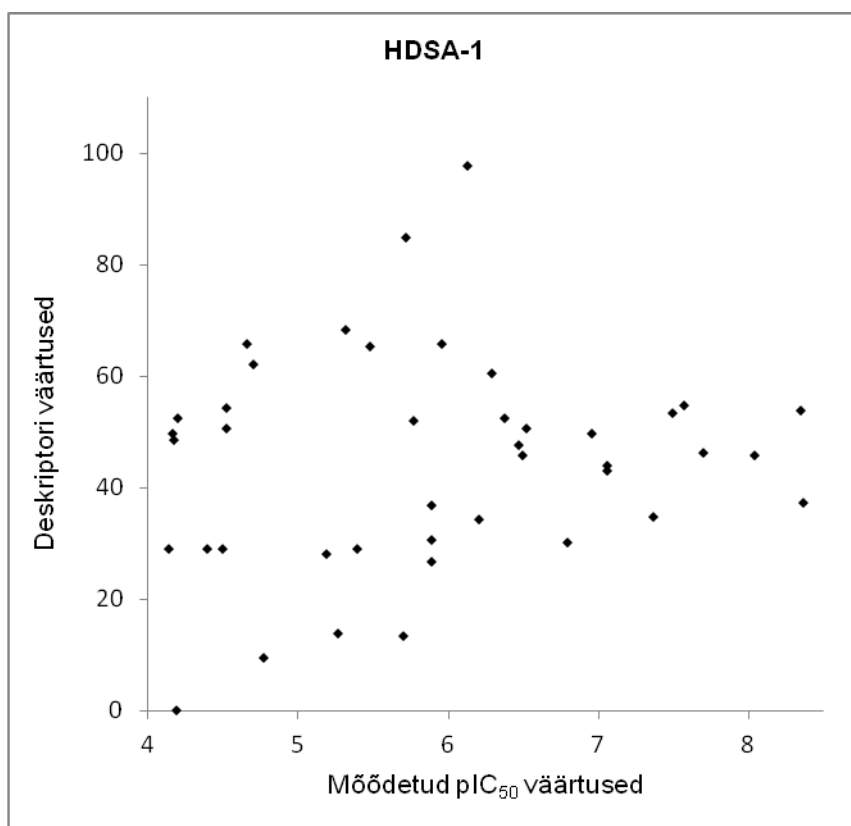
42		8,37	7,46	0,90
----	---	------	------	------

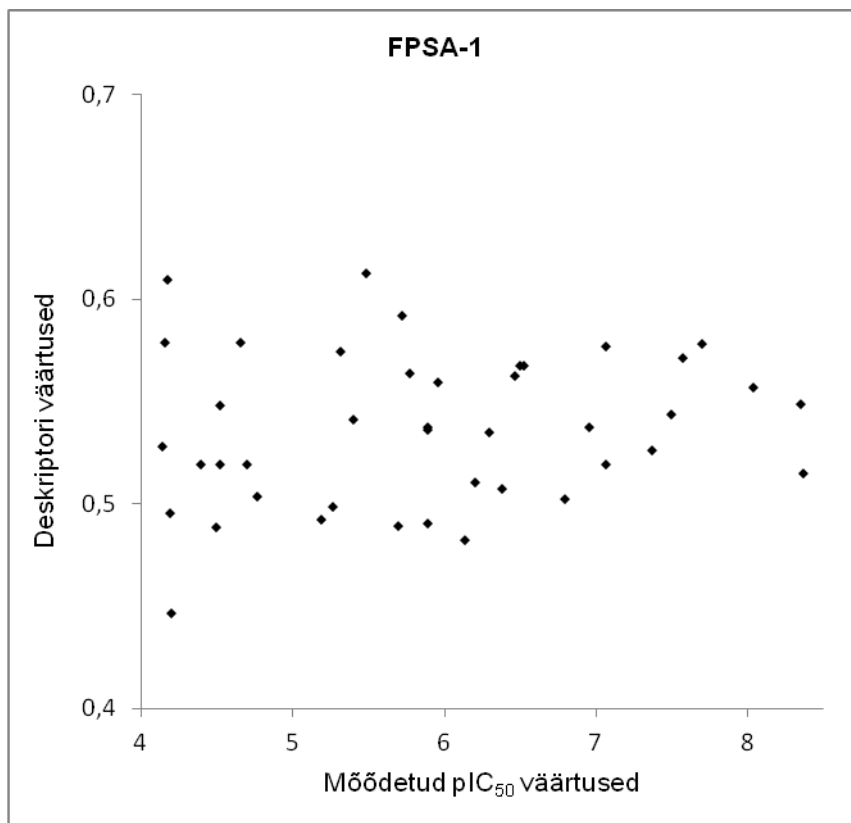
**Lisa 2.** Kolme ja viie deskriptoriga modelid.

3 deskriptorit			
#	Koefitsient	t-test	Deskriptor
0	-1,937	-2,90	Vabaliige
1	0,013	12,21	TMSA Total molecular surface area
2	-0,044	-5,81	HA dependent HDSA-1 (Zefirov PC)
3	0,031	4,28	HDSA H-donors surface area (MOPAC PC)
$N = 42; n = 3; R^2 = 0,802; R^2_{cv} = 0,762; F = 51,35; s^2 = 0,33.$			
5 deskriptorit			
0	-4,065	-3,50	Vabaliige
1	0,128	7,21	Bonding Information content (order 2)
2	-0,046	-7,46	HA dependent HDSA-1 (Zefirov PC)
3	0,012	5,18	PNSA-1 Partial negative surface area (MOPAC PC)
4	0,028	4,80	HDSA H-donors surface area (MOPAC PC)
5	-33,53	-2,94	Max partial charge (Zefirov) for atoms for atom O
$N = 42; n = 5; R^2 = 0,877; R^2_{cv} = 0,845; F = 51,20; s^2 = 0,21.$			
$N$ – ühendite arv, $n$ – deskriptorite arv, $R^2$ – korrelatsioonikoefitsient, $R^2_{cv}$ – ristvalideeritud korrelatsioonikoefitsient, $F$ – Fisheri kriteerium, $s^2$ – standardhälbe ruut.			

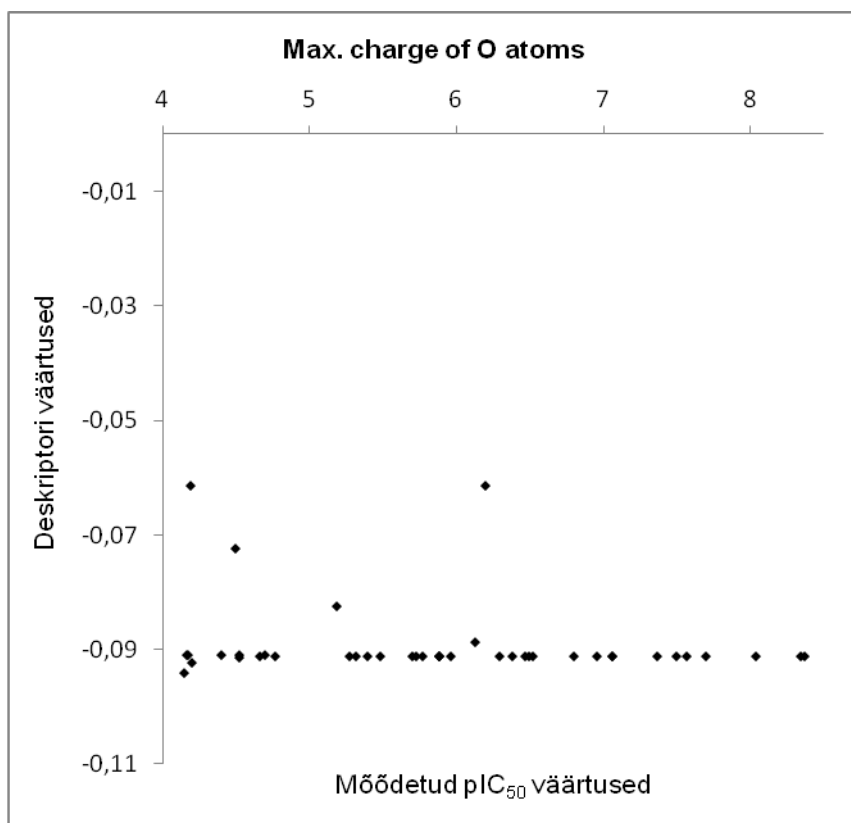
**Lisa 3.** Nelja deskriptoriga mudeli deskriptorite jaotumine.







**Lisa 4.** Viie deskriptoriga mudeli „hapniku aatomite maksimaalne osalaeng“ deskriptorijaotumine.



## **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina, Hardi Haljas  
(sünnikuupäev: 18.09.1984)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

„Inimese papilloomiviiruse HPV6 E1 helikaasi ATPasse aktiivsuse inhibeerimise modelleerimine QSAR meetodiga“,

mille juhendajad on Kaido Tamm ja Lauri Sikk,

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **29.05.2013**